L1: Entry 25 of 45 File: USPT Jan 13, 2004

DOCUMENT-IDENTIFIER: US 6678250 B1
TITLE: Method and system for monitoring and management of the performance of real-time networks

Brief Summary Text (2):
This invention relates in general to the monitoring and analysis of communication networks and more particularly, it is directed to providing a system capable of monitoring the quality of individual voice, video and other real-time connections, and also monitor and evaluate the overall performance of packet-based networks.

Brief Summary Text (4):
The Internet is a worldwide network of interconnected computers allowing users to transmit, access and obtain files from other computers and users on the network. In recent years, the Internet is becoming used more and more for new real-time applications such as allowing telephone callers to place voice telephone calls over the Internet. In addition, other real-time interactive applications such as videoconferencing allow users to conduct meetings over the network without having to physically travel to the meeting location or incur long distance communication charges.

Brief Summary Text (5):
The performance of these network applications, however, has generally disappointed users due to the vagaries of the performance and reliability of interactive communication applications over packet-based networks such as the Internet. The Internet was not originally designed for interactive communication, but rather, for the bulk transport of packets of digital data using non-interactive protocols, such as sending electronic mail ("E-mail"), File Transfer Protocol ("FTP"), and network news, i.e., USENET. Regardless, a number of real-time protocols have been designed and marketed, including RealAudio by RealNetworks, NetMeeting by Microsoft, and many others.

Brief Summary Text (6):
The performance of the network implementing these applications, however, is difficult for network operators to monitor and determine. The performance of interactive network application software operating over the Internet or other packet-based networks depends on the ability of the network to deliver digital packets of audio and video information between callers. Internet packet delivery delay and loss dynamics, however, can be extremely variable. Packet delay and loss characteristics between two callers or hosts devices may remain stationary for an hour or more, or they may change dramatically second-by-second. Different pairs of host devices communicating over the same network may also experience different network performance dynamics due to available network bandwidth and background traffic patterns. This temporal and spatial heterogeneity of the network performance makes it difficult to monitor the performance of the network. Without visibility of network performance, operators have difficulty identifying and addressing trouble areas to improve performance.

Brief Summary Text (11):
In an illustrative embodiment of the invention describing a Voice-over-IP (VoIP) network, the voice gateways transmit audio, video and other data in RTP (Real-time

Transport Protocol) streams. RTP includes a control protocol, RTCP (RTP Control Protocol), which allows session members participating in a session to exchange information related to network performance. RTCP collects statistics on the quality of the transport service between session members; i.e., remote applications communicating over the network and transmits the statistical data between the session members. The illustrative embodiment presented here utilizes RTCP to generate and transmit the relevant network performance statistics. The network performance is organized and maintained in databases and monitored to analyze the data to provide information regarding network performance.

Detailed Description Text (3):
FIG. 1 shows a simplified block diagram illustrating the high-level architecture of an exemplary packet-based network 30 ("PBN"). Analog phone calls are initiated at analog telephones 12 that establish call connections to access the network 30 through the Public Switched Network ("PSTN") 16. The PSTN call connections are terminated at modems (not shown) provided by edge devices 20, 21, 22, and 23 to interface the network 30. Preferably, the network 30 is an Internet Protocol ("IP") based digital network. This illustrative embodiment refers to an exemplary IP-based network for purposes of simplicity and clarity of explanation. Of course, the network 30 can be any of a variety of packet-based networks and interconnected digital networks including private networks, the Internet, intranets and other digital communication networks. It should also be understood that the analog telephones 12 need not be connected through the PSTN 16 to access the network 30. In addition, digital telephones 13 that connect directly to a digital network 15 such as a local area network or intranet that are capable of connecting to a PBN may also be used.

Detailed Description Text (4):
The illustrative embodiment of the network performance monitoring system is directed to a Voice-over Internet Protocol ("VoIP") application. In the context of a VoIP application, the network devices providing voice access to the network may commonly be referred to as voice gateways. In a VoIP system, the real-time voice data are transported as digital data packets. It should be understood that in other applications, the devices accessing the network 30 need not be analog telephones 12 but may be other communication and computing devices such as video cameras 17, personal computers 18, modems 19, etc.

Detailed Description Text (7):
RTP is primarily designed to provide end-to-end network transport functions suitable for real-time network applications such as a VoIP application transmitting real-time audio data over the network. RTP is more fully defined by the Internet Engineering Task Force "RTP: A Transport Protocol for Real-Time Applications" Nov. 18, 1998, ietf-avt-rtp-new-02.txt, Schulzrinne et al. ("RTP paper"), which is fully incorporated herein by reference. RTP is designed to optimize the end-system processing speed for the real-time data such as interactive voice and video data. The RTP packets are themselves transported within another protocol such as UDP (User Datagram Protocol) packets on the IP network. UDP provides multiplexing and checksum services, however, RTP may be used with a variety of different underlying network or transport protocols as known to those of skill in the art and those yet to be promulgated. A router function in the gateway 20 directs the RTP packets onto the IP network 30 that transports the packet to the destination voice gateway 23.

Detailed Description Text (9):
RTP is particularly designed to satisfy the needs of multi-participant multimedia conferences, such as multiple party video conference calls that involve many streams of audio and video transmitted to multiple callers. For example, to provide a videoconference connection with multiple callers simultaneously communicating with one another requires multiple sessions sending media streams between all the call participants. RTP is capable of supporting data transfer to multiple destinations using multicast distribution provided by the underlying network and

transport protocols. A stream of such RTP packets that are associated with a given telephone or videoconference connection is said to belong to an RTP session. The RTP session identifies the call, and session members all participate in the call to receive call information. If both audio and video media are used in a connection, the audio and video streams are transmitted as separate RTP sessions. Multi-party conference calls may thus require multiple RTP sessions, with multiple participants per session.

Detailed Description Text (12):
In addition to real-time data transmission, RTP includes a control protocol, RTCP (RTP Control Protocol), which allows session members of an on-going session to monitor and exchange information related to the network performance, as well as, providing minimal control signaling functions. RTCP is also more fully described in the RTP paper. A large part of RTCP is aimed at generating statistics on the quality of the packet transport service between session members; i.e., remote applications communicating via RTP streams. RTCP is particularly directed for generating and collecting packet delivery performance specifications on an individual, per-connection basis. RTCP, however, provides only individual connection statistics and does not provide a picture of overall network performance. The embodiment presented here utilizes RTCP to generate the relevant per-connection performance statistics and defines how they are collected and maintained in databases 14 (FIG. 1) to build a network management system capable of providing network performance information. Preferably, each gateway 20 includes a database 14 to collect and maintain network performance information generated by RTCP or other similar protocols or processes.

Detailed Description Text (22):
As seen in the embodiment of FIG. 3, a reception report block 42 includes a number data of fields containing network data including network packet delivery performance data pertaining to the call session. For example a reception report 42 preferably provides network performance statistics regarding the reception of RTP data packets received from the source gateway that the reception report 42 block corresponds to. The first field 45 of the reception report block 42 indicates the source gateway that the data in the reception report block 42 pertains to, in this example SSRC_1. Because one of the gateways in the gateway session pair will be the source of the transmission, its IP address is already contained in the IP header of the transmission. Thus, only one additional 32-bit value is required to include the IP address of the gateway at the other end of the session.

Detailed Description Text (34):
At the lowest level of the network hierarchy, gateways 69, 70, 71, 72, 73 are grouped in units referred to in this embodiment as Level_0 clusters 66, 67, 68. Level_0 clusters are comprised of a set of gateways referred to as Level_0 members. In a Level_0 cluster 66 each gateway 69, 70, 71, 72, 73 may communicate network performance data with any other gateway 69, 70, 71, 72, 73 in the Level_0 cluster 66 such that the Level_0 cluster also defines every possible gateway pair that can be formed by its members. The term Level_0 cluster pair is used to define a gateway pair 74 for two gateways 71, 73 belonging to the same Level_0 cluster 66. In this model, the case of multiple, co-located gateways is considered a single, compound gateway. The network monitor 60 for given Level_0 cluster 60 is responsible for monitoring the network conditions between all Level_0 cluster pairs in its Level_0 cluster 60. The term Level_0 cluster monitor 60 is used to define this monitor.

Detailed Description Text (39):
A feature of RTP that can be implemented in the preferred embodiment provides gateways with the ability to dynamically adjust the transmission interval between successive RTCP packets to prevent RR and SR messages from consuming or "flooding" the total bandwidth available for the gateway session members to communicate RTP data packets. In an exemplary embodiment, the session members employ an algorithm to dynamically adjust the interval between RR and SR messages from a given session

member to maintain an upper bound of 5% on the fraction of the session bandwidth consumed by RTCP traffic.

Detailed Description Text (46):
As part of the process of generating the reception report block 42, a copy of the reception report 42 could be "diverted" to a monitoring process on the gateway 20 to maintain the copy of the reception report 42. Typically, a reception report 42 is generated and sent to other peer systems during communication. In the present embodiment, the reception report is also maintained or diverted to other monitoring systems to maintain statistics for monitoring purposes.

Detailed Description Text (49):
Preferably, gateways maintain network performance data generated in reception reports in databases 14 (FIG. 1) organizing the data according to gateway pairs, the source gateway indicated in the reception block. For example, the combination of the source gateway indicated in the source field 45 (FIG. 3) of the reception report and the gateway on which the database 14 is being maintained defines a gateway pair. According to the network hierarchy previously described, the associated network monitor for the gateway pair is accordingly determined. Databases 14 are continuously updated to add the latest network performance statistics to any previous statistics for the gateway pairs. Periodically, the gateway database network performance information is transmitted to the appropriate network monitor associated with the gateway pair. A suggest period for transmitting updated database information is three minutes, the average length of a voice call. Preferably, statistics are maintained for the period such that at the end of each period, each updated quantity represents a time average over the period. After transmission of updated statistics, the statistics are reset in preparation for accumulation during the next period.

Detailed Description Text (64):
The present embodiment preferably includes logic to implement the described methods in software modules as a set of computer executable software instructions. The Computer Processing Unit ("CPU") or microprocessor implements the logic that controls the operation of the channel card. The microprocessor executes software that can be programmed by those of skill in the art to provide the described functionality. The software can be represent as a sequence of binary bits maintained on a computer readable medium including magnetic disks, optical disks, organic disks, and any other volatile or (e.g., Random Access memory ("RAM")) non-volatile firmware (e.g., Read Only Memory ("ROM")) storage system readable by the CPU. The memory locations where data bits are maintained also include physical locations that have particular electrical, magnetic, optical, or organic properties corresponding to the stored data bits. The software instructions are executed as data bits by the CPU with a memory system causing a transformation of the electrical signal representation, and the maintenance of data bits at memory locations in the memory system to thereby reconfigure or otherwise alter the unit's operation. The executable software code may implement, for example, the methods described in further detail below.